Okasha, S., & Paternotte, C. R. (2012). Group adaptation, formal Darwinism and contextual analysis. Journal of Evolutionary Biology, 25(6), 1127-1139. 10.1111/j.1420-9101.2012.02501.x

Title: Group Adaptation, Formal Darwinism and Contextual Analysis

Authors: Samir Okasha, Cedric Paternotte
(Order of authors is alphabetical. Both authors contributed equally to the research and writing of the paper.)

Affiliations: University of Bristol

Running Title: Group Adaptation, Formal Darwinism and Contextual Analysis

Corresponding Author: Samir Okasha, Department of Philosophy, 9 Woodland Road, University of Bristol, BS8 1TB. Tel + 44 117 9287829 Fax + 44 117 954 6001. Email: Samir.Okasha@bristol.ac.uk

Abstract

We consider the question: under what circumstances can the concept of adaptation be applied to groups, rather than individuals? Gardner and Grafen (2009) develop a novel approach to this question, building on Grafen's 'formal Darwinism' project, which defines adaptation in terms of links between evolutionary dynamics and optimization. They conclude that only clonal groups, and to a lesser extent groups in which reproductive competition is repressed, can be considered as adaptive units. We re-examine the conditions under which the selection-optimization links hold at the group level. We focus on an important distinction between two ways of understanding the links, which have different implications regarding group adaptationism. We show how the formal Darwinism approach can be reconciled with G.C. Williams' famous analysis of group adaptation, and we consider the relationships between group adaptation, the Price equation approach to multi-level selection, and the alternative approach based on contextual analysis.

Keywords: adaptation, group adaptation, superorganism, Price's equation, optimality, contextual analysis, G. C. Williams, formal Darwinism, group selection

Group Adaptation, Formal Darwinism and Contextual Analysis

1 Introduction

Evolutionary biologists usually apply the concept of adaptation to individual organisms. However it has long been recognised that in principle, groups might also exhibit adaptations. The idea of group adaptation, and the associated concept of a 'superorganism', were famously criticised by G.C. Williams (1966), but have since been revived by proponents of 'multi-level selection' (Sober & Wilson 1998; Seeley 1989, 1997; Hölldobler & Wilson 2009). Progress on this topic has been hampered by unclarity about how exactly 'group adaptation' should be defined, how it relates to 'group selection', and the conditions under which it can evolve. Gardner and Grafen (2009) (hereafter G&G) make a remarkable contribution by bringing mathematical precision to these issues, with striking results. They do this by applying Grafen's 'formal Darwinism' project, (Grafen 2002, 2006, 2008), which provides a general framework for understanding the concept of adaptation, to groups.

Our aim here is to take further the analysis of group adaptation, using a similar methodology to G&G. We recognise the merits of making the concept of group adaptation precise, and share their view that the formal Darwinism project offers the best way to do this. However, G&G's analysis leaves open a number of issues. In particular, it is unclear how the concept of group adaptation they articulate relates to G.C. Williams' (1966) well-known analysis of the concept (cf. Sober and Wilson 2011).

Our discussion falls into three parts. Firstly, we explore a subtle difference between two ways of defining adaptation using the formal Darwinism machinery, one used by G&G, the other by Grafen in his earlier papers. The two defintions have different implications in general; and as applied to groups, they differ on whether clonality, or repression of within-group competition, represent the clearest cases of group adaptation.

Secondly, we study how the formal Darwinism approach can be reconciled with G.C. Williams' distinction between 'group adaptation' and 'fortuitous group benefit'. The former refers to a group feature that evolved *because* it benefits the group, the latter to a group feature that happens to benefit the group but did not evolve for that reason. (Thus Williams famously contrasted a 'herd of fleet deer' with a 'fleet herd of deer'.) Many biologists regard this dis-

tinction as crucial, so it is of some interest to see whether the formal Darwinism approach to group adaptation can accommodate it.

Thirdly and relatedly, we consider the relation between the 'Price equation' approach to multi-level selection and the alternative approach based on 'contextual analysis'. These approaches constitute alternative ways of partitioning the total evolutionary change in a structured population into components corresponding to distinct levels of selection. G&G say that their analysis 'has identified Price's between-group selection as the driver of group adaptation', and thus favour the Price approach (p. 667). We show that the contextual approach can also supply a formal definition of group adaptation.

2 The 'maximizing agent' analogy

G&G's analysis of group adaptation draws on Grafen's 'Formal Darwinism' project, which aims to connect optimization and natural selection in a precise way, thus formally justifying the intuitive idea that selection leads to organismic design (Grafen 2002, 2006, 2008). Grafen's approach is to use a fully explicit definition of optimization, then to prove links between optimality and evolutionary dynamics. The notion of optimization is captured by an 'objective function' that maps an agent's phenotype to its 'fitness' (for some measure of fitness); if an agent achieves the maximum value of this function, they are said to 'behave optimally'. The links state logical connections between the optimality or otherwise of agents' behaviour and the operation of natural selection.

In Grafen's original papers, the 'agents' are taken to be individual organisms; this is natural because individuals are usually treated as the bearers of adaptations in biology. With this interpretation, the links capture the sense in which natural selection leads individuals to be adaptive units, just as Darwin originally argued. G&G investigate what happens when the 'agents' in Grafen's analysis are instead taken to be *groups*; their aim is to see whether, and in what circumstances, whole groups can legitimately be considered as adaptive units, or 'maximizing agents'. Their main conclusion is that these circumstances are relatively rare, because the required links between optimality and natural selection only hold under fairly stringent conditions.

To understand G&G's argument, the optimality / selection links must be laid out explicitly (see Table 1). The first link says that if all agents behave optimally, there is no 'scope' for selection, i.e. no gene will change in (expected) frequency. This makes good sense: if all agents achieve maximum fitness, the fitness variance in the population is zero, so no selection will occur. The second link says that if all agents behave optimally, there is no 'potential' for positive selection, which means that no introduced mutant will spread. This also makes sense: if all agents achieve maximum fitness, then no mutant can do better. The third link says that if all agents behave suboptimally, but equally so, there is no scope for selection. Again this makes sense, given that selection requires variance in fitness. The fourth link says that if all agents behave suboptimally, but equally so, then there *is* potential for positive selection. This is also intuitive,

1. If all agents behave optimally, there is no scope for selection.
2. If all agents behave optimally, there is no potential for positive selection.
3. If all agents behave suboptimally, but equally so, there is no scope for selection.
4. If all agents behave suboptimally, but equally so, then there is potential for positive selection.
5. If agents vary in their optimality, then there is scope for selection, and the change in the frequency of any gene is given by the covariance between the frequency of that gene in an agent and the agent's relative fitness.

Table 1: The selection / optimality links

since a mutant phenotype which achieves a higher fitness than the incumbents will spread in the population.

The fifth link is slightly different, in that it describes what will happen if the agents vary in their optimality. The link says that if agents vary in their optimality, then there *is* scope for selection, and the change in the frequency of any gene is given by the covariance between the frequency of that gene in an agent and the agent's relative fitness. The first part of this is intuitive – non-zero variance in fitness implies that natural selection can operate; the second part follows from the Price equation with the second term set to zero, described below. A sixth link is discussed by G&G, but we do not treat it separately here as it is a logical consequence of links four and five taken together (as they note in their Appendix.)

These links may seem obvious, but as Grafen (2002) points out, that is only because many biologists simply take for granted that selection leads to optimization. And in fact, the assumptions that must be made, and the definition of 'fitness' that must be used, in order for the links to be proved are non-trivial matters. For example, when the agents are individuals, the absence of mutation and gametic selection must be assumed to prove the links; and depending on whether the individuals socially interact, different definitions of 'fitness' must be used (Grafen 2006). So the project is highly non-trivial.

When the agents are individuals, Grafen speaks of an 'individual as maximising agent' (IMA) analogy, to capture the idea that individuals behave like economically rational agents, attempting to maximise the value of their objective function. If all five links hold, the IMA analogy is closely tied to the action of natural selection. It is then legitimate to treat individuals as adaptive units, Grafen argues, and to regard natural selection as acting to optimize each individual's phenotype. But where the links do not hold, there is no justification for employing the concept of individual adaptation.

G&G apply a similar logic to groups, by developing a 'group as maximising agent' (GMA) analogy. They study the conditions under which the five links hold, with 'agents' understood as groups. These conditions then determine when talk of group adaptationism is valid, i.e. when it is legitimate to regard groups as adaptive units, and natural selection as acting to optimize

the group's phenotype. So for G&G, the validity of group adaptationism thus depends on whether the selection / optimality links hold, where groups are the optimizing agents. This yields an understanding of group adaptationism that is both conceptually clear and mathematically precise.

3 The selection / optimality links: 'actual' versus 'possible' definitions

For the selection / optimality links to be formally proved, they need to be expressed mathematically. 'Optimality' is defined as maximization of the objective function; 'scope for selection' and 'potential for positive selection' are expressed in terms of the evolutionary change in what Grafen (1985) calls 'p-scores'. Formally a p-score is simply a function from the set of individuals in the population to $\mathbb{R}$. In the simplest case a p-score is an indicator function for a particular allele, indicating the frequency of the allele within an individual (= 0, 1/2 or 1 for diploids); the average of this p-score over individuals is then the frequency of the allele in the population. Any weighted sum of such indicator functions also counts as a p-score, is why a p-score can assume any real value. (These weighted sums represent breeding values of phenotypic traits; see Grafen (1985, 2002, 2008) for a full explanation of p-scores.)

In Grafen's (2002) discussion of the IMA analogy, and in Grafen (2006), he considers the set of all possible p-scores in a population, i.e. all functions from the set of individuals to $\mathbb{R}$, irrespective of whether these functions indicate the frequency of an allele actually found in the population (or a weighted sum of such functions). So even if two individuals are genotypically identical, there is still some possible p-score for which they differ. Grafen (2002) then defines 'no scope for selection', an expression that occurs in links 1 and 3, as 'no expected change in population-wide average p-score, for *any possible* p-score.' Let us call this definition 'no scope for selection (possible)'.

Grafen's definition of 'no scope for selection' may seem odd; surely it would be more natural to define it in terms of actual p-scores, rather than all possible p-scores? An 'actual p-score' may be defined as an indicator function for an allele that is actually present in the population (or a weighted sum of such functions). So for a given population, the set of actual p-scores is a proper subset of the set of all possible p-scores. If 'no scope for selection' were defined in terms of actual p-scores, it would mean that there whenever there is no expected genetic change in a population, there is no scope for selection, and vice-versa. Let us call this definition 'no scope for selection (actual)'.

The biological meaning of the condition 'no scope for selection (actual)' is obvious, but what about 'no scope for selection (possible)'? In effect, the latter condition means that no allele actually present in the population will change in expected frequency *and* that no *neutral* mutations can change in expected frequency. (By contrast, 'no potential for positive selection' concerns the fate of non-neutral mutations.) Conversely, if there *is* 'scope for selection (possible)' in

a population, this means that the fitness distribution is such that, if the requisite genetic variation were present, there would be expected gene frequency change. So although the condition 'no scope for selection (possible)' seems odd at first sight, referring as it does to non-actual p-scores, it can be given a reasonable biological interpretation.

Moreover, the 'possible' defintion is crucial to Grafen's project. To see why, consider link 5 – which says that if agents vary in optimality then there *is* scope for selection. Suppose a population of individuals does exhibit variance in optimality (fitness), but is in population-genetic equilibrium. This could be for a number of reasons, e.g. overdominance. For example, suppose that individual fitness depends exclusively on genotype at a single heterotic locus; assume that AA and BB individuals are non-viable, while ABs are viable. So at equilibrium, the individuals do vary in optimality. However, at the locus in question there will be no evolutionary change; and we may assume that at every other locus, all individuals are genotypically identical. So no allele present in the population will change in expected frequency; thus there will be no expected change in any actual p-score. However, there does exist some *possible* p-score, e.g. whose value is positively correlated with individual fitness, which will change in frequency. So for link 5 to be true in the IMA case, 'no scope for selection' has to be defined with reference to all possible p-scores, rather than just actual p-scores.

It might be argued that the use of 'all possible' p-scores, in the definition of 'scope for selection', is unnecessary, as a referee suggests, for the following reason. In the overdominance example. there are exactly two possibilities: either all individuals are genetically identical at all loci over than the overdominant locus (case A), or this is not so (case B). Both possibilities are consistent with the model assumptions. If we do not know whether case A or case B obtains, then for all we know, there may be an allele actually present in the population which will change in expected frequency. Since we cannot rule this out, in this sense there is 'scope for selection' based solely on change in actual p-scores. The problem with this reasoning is that it makes the existence or otherwise of 'scope for selection' dependent on our knowledge, rather than a matter of objective fact. We regards this as undesirable, since the holding of the selection / optimality links, and thus the validity of adaptationism, would then become knowledge-relative too. To avoid these untoward consequences, one must allow that there is 'scope for selection' in both cases A and B above, which is precisely what Grafen (2002) achieves by defining 'scope' in terms of all possible p-scores. So the distinction between 'actual' and 'possible' p-scores is necessary.

In the IMA case, it is easy to see that link 5 is the only link that could not be proved using the weaker 'actual' definition of 'no scope for selection', under the assumptions of no mutation or gametic selection. (The expression 'scope for selection' does not occur in links 2 and 4, while links 1 and 3 must hold on the 'actual' definition whenever they hold on the 'possible' definition.) However we show below that in the GMA case, link 5 can hold even on the 'actual' definition of 'no scope for selection', in certain special cases; and moreover, link 5 can fail to hold even on the 'possible' definition, in certain other cases.

If we accept the basic logic of the formal Darwinism approach – that adap-

tationism is defined by the five selection / optimality links holding – then the distinction between the 'actual' and 'possible' definitions of 'scope for selection' gives rise to two subtly different forms of adaptationism. It is an open question which is better. Queller and Strassman (2009) have recently argued that whether some entity is a 'unit of adaptation' depends on the extent of actual, not possible, selective processes within that entity. We do not take a stand on this issue here. In what follows we do not endorse either of the two definitions of 'scope for selection' as objectively correct, but rather explore the logical consequences of both.

4 Groups as Adaptive Units

Gardner and Grafen (2009) consider a model of evolution in a structured population. There are M groups each containing N individuals. Each individual has a genotype, a phenotype, and a reproductive success value. Each group has a 'group genotype', which is an unordered list of the genotypes of its constituent individuals; group genotype determines group phenotype, which determines group reproductive success. As before, a p-score is a function from the set of MxN individuals to $\mathbb{R}$.

The evolutionary change in any p-score is described by the change in the average p-score in the population over one generation, which we denote $\Delta\overline{p}$. Gardner and Grafen treat $\Delta\overline{p}$ as a random variable, in order to model uncertainty, and focus on its expected value. Explicitly incorporating uncertainty allows them to handle many biological complexities; however these are not relevant for our purposes, so to keep the analysis simple we ignore uncertainty and talk about the actual change. This is strictly for simplicity; the expected change is what really matters, and our results could easily be formulated in such terms.

Assuming no gametic selection or mutation, $\Delta\overline{p}$ is given by the simplest form of the Price equation:

$$\overline{w}\Delta\overline{p} = Cov_{I\text{x}J}(w_{ij}, p_{ij}) \tag{1}$$

where p_{ij} and w_{ij} are the p-score and the reproductive success of the j^{th} individual in the i^{th} group respectively; $I = \{1, \ldots, M\}$ is the set of group indices and $J = \{1, \ldots, N\}$ the set of individual indices; and $\overline{w}$ is average reproductive success in the population.

As is well-known, equation (1) can be expanded into a 'multi-level' format, by partitioning the total covariance between individual p-score and individual reproductive success into between-group and within-group components, yielding the result first obtained by Price (1972):

$$\overline{w}\Delta\overline{p} = Cov_I(w_i, p_i) + E_I[Cov_J(p_{ij}, w_{ij})] \tag{2}$$

where w_i is the average reproductive success of the i^{th} group, p_i the average p-score of the i^{th} group. The first term on the RHS is the covariance between a group's average p-score and its group reproductive success; the second term

is the average, or expectation, across groups of the within-group covariance between individual p-score and individual reproductive success. Equation (2) is often regarded as decomposing the total change into components corresponding to the effects of 'group selection' and 'individual selection' respectively. This interpretation is standard in the literature on multi-level selection, though it is not the only way that these contested terms have been defined.

It is a familiar point that substantial within-group selection may undermine group functionality, thus preventing the group from behaving as an adaptive unit (Buss 1987, Maynard Smith & Szathmary 1995, Frank 2003). G&G thus consider two models in which within-group selection is completely absent, which should consitute a 'best-case scenario' for group adaptationism. The first involves purely clonal groups; the second involves non-clonal groups with full repression of competition, i.e. no within-group variance in fitness. (By this G&G mean no within-group variance in *expected*, rather than realized, fitness – which means that the existence of reproductive division of labour in a group is fully compatible with zero within-group variance in fitness. This is one place where the distinction between realized and expected fitness, which we are ignoring for simplicity, is important.)

If there is no within-group selection on a given p-score, the second RHS term of equation (2) will be zero, in which case it reduces to

$$\overline{w}\Delta\overline{p} = Cov_I(w_i, p_i) \tag{3}$$

Clearly, equation (3) will apply to any p-score that shows no within-group variance. So in the clonal groups model, equation (3) will apply to all *actual* p-scores. Similarly, equation (3) will apply whenever there is no within-group variance in fitness, as in the repression of competition model. Both models imply that for each group, the within-group covariance between fitness and p-score is zero, and thus the average of this covariance across groups is also zero.

G&G then claim that in both of these models, the links between the GMA analogy and gene frequency change do obtain (with one proviso), so group adaptationism is valid. This is the central positive claim of their paper. The reason the links hold in these models, they claim, is that the assumption of no within-group selection renders equation (3) applicable, which in turn allows the five links to be proved, with the objective function taken to be group fitness, i.e. the average fitness of the individuals in the groups. By contrast, when within-group selection is not assumed absent, so the full Price equation (2) must be applied, none of the links can be proved, so it is not legitimate to regard groups as adaptive units.

The proviso concerns link 4 in the repression of competition model (which says that if all agents behave equally suboptimally, then at least one mutant can spread). This need not be true, G&G argue, because although an improved group phenotype is possible at the suboptimal equilibrium, "there is no guarantee that the corresponding genetic variants will arrange themselves together in groups in such a way as to give rise to the desired group phenotype" (p. 665). In the clonal case this problem doesn't arise, since any group phenotype can

be produced by a single genetic variant. So they regard talk of group adaptation as fully justifiable in the clonal case, but only partly justifiable in the repression-of-competition case.

The significance of this consideration is debatable, since a parallel problem arguably applies at the individual level too. In Grafen (2002), where link 4 is proved for individuals, it is simply assumed that any non-resident phenotype can be produced by a genetic variant – even though this may require several simultaneous mutations at different loci. A parallel assumption could be made in the group case, i.e. that any non-resident group phenotype will be produced by mutation, even if this requires several individuals to mutate simultaneously – in which case link 4 would be true. It may be that the required assumption is less plausible in the group than the individual case, but this is an empirical matter. Therefore, we are inclined to regard link 4 as equally defensible, in principle, in both the repression-of-competition and clonal models. But nothing in what follows turns on this.

5 Clonality versus Repression of Competition

Aside from the proviso concerning link 4, G&G treat clonal groups and competitively repressed groups on a par. However, there is actually a logical difference between them with respect to links 1 and 3. For simplicity we focus on link 1, which to recall says that if all groups are optimal, then there is no scope for selection. Recall the distinction between the 'actual' and 'possible' definitions of 'no scope for selection' from section 3. If we adopt Grafen's original 'possible' definition, it turns out that link 1 is true in the repression of competition model but *not* in the clonal groups model.

Repression of competition implies that there is no within-group variance in fitness. (We do not take this condition to $define$ repression of competition, for it is possible that within-group fitnesses may be equal anyway. Repression is a causal mechanism for bringing this about.) The absence of within-group variance in fitness can be expressed by $Var_J(w_{ij}) = 0$ for all groups i. This implies that for every possible p-score, $Cov_J(w_{ij}, p_{ij}) = 0$ in each group i, which implies that equation (3) above describes the evolutionary dynamics of each p-score. Link 1 then follows immediately; since if all groups are optimal then there is no variance in group fitness, so equation (3) tells us that $\Delta p = 0$ for every possible p-score.

Now consider clonality. Note firstly that clonal groups cannot be defined as $Var_J(p_{ij}) = 0$ for all possible p-scores and all groups i, i.e. no within-group variance in any possible p-score in any group. For this condition is logically unsatisfiable, given that the set of possible p-scores is the set of all functions from the set of individuals to $\mathbb{R}$. That groups are clonal means the absence of within-group variance in any *actual* p-score. But there will be many possible p-scores that do show within-group variance, even if the groups are clonal.

This means that the condition $Cov_J(w_{ij}, p_{ij}) = 0$ in each group i does not hold for every possible p-score in the clonal group model, unlike in the repression

of competition model (see Appendix 1). Of course, even if that condition does not hold for a given p-score, equation (3) could still apply to that p-score if the weaker condition $E_I[Cov_J(w_{ij}, p_{ij})] = 0$ holds, i.e. the average over groups of the within-group covariances is zero. However, this latter condition cannot hold true for all p-scores, unless within-group fitnesses are equal. (See Appendix 2, Propostion 2, for proof.)

This means that on the 'possible' definition of scope for selection, link 1 only holds in the clonal groups model if there is no variance in within-group fitnesses in any group. Consider a case where the groups are clonal but within-group fitnesses do vary. In this case, it is not true that if all groups are optimal, $\Delta p = 0$ for every possible p-score. It will always be possible to find a p-score for which the condition $E_I[Cov_J(w_{ij}, p_{ij})] = 0$ does not hold, and for which Δp will be non-zero. So even if all groups are optimal, there will always be scope for selection unless within-group fitnesses are equal in each group. In fact, the absence of within-group variance in fitness turns out to be necessary and sufficient for all the links to hold, as we show in section 7.

In a clonal groups scenario, it is of course possible that within-group fitnesses will be equal. This will be so if individual fitness depends only on individual genotype. But this need not be true. There are various reasons why the members of a clonal group may differ in fitness (aside from chance), e.g. they may receive different amounts of social help. It might still be argued that their *expected* fitnesses will be equal, but this depends on how exactly the states of the world, over which the expectation is taken, are defined. In any case, even if it is assumed that clonal group mates have the same expected fitness, in which case link 1 will hold on the 'possible' definition of scope for selection, it is important to realise that it is not clonality but rather the absence of within-group fitness variance that is responsible for the link holding.

Since G&G hold that there can be no scope for selection within clonal groups, in virtue of the clonality, it is clear that they are employing the 'actual' definition of 'scope for selection', on which links 1 and 3 do indeed hold for clonal groups. This definition is perfectly reasonable, but as we saw in section 3, adopting it complicates the formal Darwinism approach to individual adapation, as it makes link 5 logically stronger and thus harder to satisfy. In the group case, adopting the 'actual' definition of scope for selection similarly strengthens link 5; as a result, repression of within-group competition no longer suffices for link 5 to hold, but clonality does.

To understand this, consider the following example. A population contains asexual individuals of two genotypes, A and B, living in groups of size $N = 4$. Groups are competitively repressed, so within each group all individuals have the same fitness. The population contains exactly three types of group: $(AAAA)$, $(BBBB)$ and $(AABB)$; the group fitness function is non-linear, and is such that $w(AABB) > w(BBBB) = w(AAAA)$. (This is a group-level analogue of over-dominance.) As a result, the population is in equilibrium – no gene will change in frequency – but the groups do vary in fitness (optimality). So link 5, which says that if the groups vary in fitness then there is scope for selection, need not be true for competitively repressed groups under the 'actual' definition

	'Possible' definition	'Actual' definition
Individual	Links 1–5 ✓	Links 1–4 ✓, link 5 x
Group–Clonality	Links 2,4,5 ✓, links 1,3 x *	Links 1–5 ✓ **
Group–Repression	Links 1–5 ✓	Links 1–4 ✓, link 5 x

Table 2: Conditions under which the links hold.
* Links 1 and 3 will hold if clonal group mates have identical fitness.
** Link 5 will fail if group fitness does not depend only on group genotype.

of scope for selection.

This counterexample to link 5 depends essentially on the groups being non-clonal. This is because for there to be a polymorphic equilbrium with fitness differences between groups, it is essential that some groups contain individuals of different genotypes, given that group fitness depends only on group genotype. Therefore, in the clonal groups model, adopting the 'actual' defintion of scope for selection does not allow a counterexample to link 5 to be constructed. (Note however that if the assumption that group fitness depends only on group genotype were relaxed, then link 5 would fail even in the clonal case.)

The upshot is that depending on whether we use the 'possible' or the 'actual' definition of scope for selection, the selection / optimality links will hold true under different conditions. These differences are summarized in Table 2, for both the individual and group models, under the standard assumptions of no mutation and no gametic selection.

What should we conclude from this? In one respect competitively repressed groups constitute the better case for group adaptationism, but in another respect clonal groups do. If we adopt the 'possible' definition of scope for selection, then repression of competition guarantees that links 1-5 hold, but clonality does not. Some biologists would regard this as welcome result. Queller and Strassman (2009) have argued that a clonal group should not automatically be regarded as a superorganism, if it shows no functional integration and no social interaction among its constituent individuals; see also Ratnieks and Reeve (1992). In a similar vein, Michod (1999) argues that true higher-level individuals (or superorganisms) must possess mechanisms for conflict suppression. By these authors' lights, an analysis of group adaptation that privileges repression of competition is independently desirable.

However if we adopt the 'actual' definition of scope for selection, then clonal groups emerge as the better candidate for the superorganism mantle. On this defintion, link 5 fails in the repression of competition model but holds in the clonal groups model (so long as group fitness is assumed to be a function of group genotype). This consideration provides a possible basis, over and above the argument given by G&G in relation to link 4, for treating clonality as the 'best case' for group adaptationism.

The dichotomy between clonality and repression of within-group competition, as means for unfiying the evolutionary interests of group members, has relevance in relation to 'major evolutionary transitions'. Multi-cellular organ-

isms typically employ both means; their constituent cells are usually genetically identical, and the fairness of meiosis serves to repress reproductive competition between the genes within a single genome, in sexual species. Indeed the assumption of fair meiosis, i.e. the absence of gametic selection, is precisely why links 1 and 3 hold true in the individual model of Grafen (2002), on the 'possible' definition of scope for selection.

We take no stand on whether the 'actual' or 'possible' defintion of scope for selection is preferable, nor therefore on whether clonality or repression of competition constitutes the better case of group adaptationism. Our aim has been to explore the logic of formal Darwinism as applied to groups, under both definitions. However in what follows we focus on Grafen's original 'possible' definition, not because we think it is intrinsically superior, but because it allows us to find necessary and sufficient conditions, that are biologically meaningful, for links 1-5 to hold.

6 Group Adaptation versus Fortuitous Group Benefit

In *Adaptation and Natural Selection*, G.C. Williams (1966) distinguished between 'group adaptation' and 'fortuitous group benefit', as part of his celebrated attack on group selectionism. The former refers to a group feature that evolved *because* it benefits the group, the latter to a group feature that happens to benefit the group but did not evolve for that reason. So on Williams' view, whether a particular feature constitutes a group adaptation depends crucially on its causal history. A clonal group of non-social aphids, or of some marine invertebrate species, would not count as group adaptation by Williams's lights, for the members of such groups engage in no social behaviour, and the groups exhibit little or no functional organization. If some such groups do better than others, this is most likely a side-effect of differences in individual adaptedness.

How does Williams' influential concept of group adaptation relate to the concept defined by the formal Darwinism approach of G&G? The concepts are clearly different; G&G hold that group adaptationism applies to any clonal group, while Williams explicitly rules out some clonal groups. From Williams' viewpoint, the five selection / optimality links which G&G take to define group adaptation could hold 'for the wrong reason', i.e. as a side-effect of individual-level processes. This would be so in a case in which there is no within-group variation in fitness, and the individuals in each group engage in no social behaviour. Williams would categorize this as fortuitous group benefit, not group adaptation.

This difference between G&G's and Williams' concepts may seem puzzling, since Williams' point was precisely that a trait only counts as group adaptation if it has evolved by a process of group-level selection; and G&G define 'group selection' as "that part of gene-frequency change that is responsible for group adaptation" (p.667). So where does the difference stem from?

The answer is that G&G identify 'group selection' with the between-group component of the multi-level Price equation, i.e. the term $Cov_I(w_i, p_i)$ in equation (2); while a proponent of Williams' view must reject this defintion. As many authors have pointed out, the multi-level Price equation is arguably a flawed way to decompose the total change into components corresponding to distinct levels of selection (Grafen 1984, Nunney 1985, Heisler & Damuth 1987, Goodnight et al. 1992; Okasha 2004, 2006). The basic problem is that the covariance between group p-score and group fitness may be positive even in the absence of any causal relation between these variables; groups with a high p-score may be fitter, simply because they contain a higher proportion of intrinsically fit individuals, even if there is no group effect on fitness, and no social behaviour. Arguably it is unhelpful to speak of 'group selection' in such a circumstance; individual selection is responsible for the entirety of the evolutionary change. This is a close corollary of Williams' point that group-beneficial features may arise as a side-effect of individual selection.

If we accept that group and individual selection should not be identified with the components of the multi-level Price equation, then what decomposition of the evolutionary change should be used to define them? One promising approach is to use 'contextual analysis', a form of multiple regression analysis (cf. Heisler & Damuth 1987). This permits a solution to the problem that besets the Price approach (i.e. the multi-level decomposition in equation (2)), by isolating the effect of a trait on group fitness once individual effects have been stripped away. The total change can still be partitioned into two components, corresponding to the two levels of selection. The crucial difference with the Price approach is that contextual analysis only identifies a component of group selection when there is a 'group effect' on individual fitness. The method is described fully in section 7.

G&G discuss contextual analysis, but appear to regard the distinction between the Price and contextual approaches as merely semantic. Clearly it is a semantic matter how we use the expressions 'group selection' and 'individual selection', but the question of whether the causal action of natural selection operates at the individual or group level is non-semantic. We accept G&G's idea that group selection should be defined as the part of gene-frequency change that is responsible for group adaptation, but we show in section 8 that this does not discriminate between the Price and the contextual definitions of group selection.

Some biologists might simply reject Williams' distinction outright, and thus reject the idea that the selection / optimality links might hold 'for the wrong reason'. Anyone doing this would naturally accept the Price decomposition, and G&G's analysis. However many authors, ourselves included, regard Williams' distinction between group adaptation and fortuitous group benefit as important. We show in section 9 that accepting this view does not mean abandoning the formal Darwinism approach altogether.

7 Price's Equation versus Contextual Analysis

Contextual analysis treats every individual in the population as having two trait values, an individual p-score and the p-score of the group it belongs to. The key question is then whether there is an association between fitness and group p-score that does *not* result from an association between fitness and individual p-score. This is assessed with a linear regression model:

$$w_{ij} = \beta_1 p_{ij} + \beta_2 p_i + e_{ij} \tag{4}$$

where β_1 is the partial regression of individual fitness on individual p-score, controlling for group p-score; β_2 is the partial regression of individual fitness on group p-score, controlling for individual p-score; and e_{ij} is the residual whose variance is to be minimized. Therefore β_2 is the change in individual fitness that would result if the group p-score of an individual of fixed p-score were changed by one unit – it measures the extent to which differences in group p-score explain differences in individual fitness, holding individual p-score constant.

If β_2 is zero, this means that an individual's fitness depends only on its own p-score, so any covariance between group p-score and fitness is a side-effect of individual selection. Intuitively this means that individual selection is the only force affecting the evolution of the p-score in the population – at least if we follow Grafen (1984) in defining 'individual selection' in terms of an action's 'effects on the actor's number of offspring alone' (p.83-4). This means that for group selection to operate, β_2 must be non-zero.

It is natural to interpret β_1 and β_2 as measures of the direct causal influence of individual p-score and group p-score, respectively, on individual fitness. However this interpretation is only valid if the true dependence of w_{ij} on p_{ij} and p_i is linear (as for example in a linear public goods game). Of course, even if the true dependence is non-linear, it is possible to apply equation (4); but in that case the partial regression coefficients cannot be construed as measures of direct causal influence.

Using contextual analysis, we can partition the evolutionary change in p-score into two components, corresponding to individual and group selection as understood here. To do this, we simply substitute equation (4) into (1). After simplifying, this gives:

$$\begin{aligned} \overline{w}\Delta\overline{p} &= \beta_2 Cov_{I\text{x}J}(p_{ij}, p_i) + \beta_1 Var_{I\text{x}J}(p_{ij}) \\ &= \beta_2 Var_I(p_i) + \beta_1 Var_{I\text{x}J}(p_{ij}) \end{aligned} \tag{5}$$

Equation (5) constitutes an alternative to the Price decomposition given in equation (2), which to recall is:

$$\overline{w}\Delta\overline{p} = Cov_I(w_i, p_i) + E_I[Cov_J(p_{ij}, w_{ij})] \tag{2}$$

Note that equations (2) and (5) are both true; but they slice up the total change in different ways. Which equation we favour depends on whether we think 'individual selection' and 'group selection' should be understood as

within-group and between-group selection, or as selection on the component of individual fitness that is due to differences in individual p-score, and to differences in group p-score.

The contextual approach to multi-level selection, enshrined in (5), tallies neatly with Williams' point that 'fortuitous group benefit' and group adaptation are different matters. In cases of fortuitous group benefit, a trait (or p-score) that is individually advantageous leads to an incidental benefit for the group; so group p-score will covary positively with the fitness of both individuals and groups. But this association goes away if we control for individual p-score, as it alone affects individual fitness; therefore β_2 is zero. On the contextual approach, the evolutionary change is then solely attributable to individual selection, whereas the Price approach wrongly detects a component of group selection.

One limitation of the contextual approach is that if a particular p-score shows no variation within groups, then the partial regression coefficients β_1 and β_2 are undefined. This is because the absence of within-group variance in p-score means that an individual's p-score and the p-score of their group are perfectly colinear – so it is impossible to compare the difference in fitness of two individuals with the same group p-score but different individual p-scores, and vice-versa. Although equation (5) cannot be applied in such a circumstance, it still makes sense to ask whether there is a direct causal link between individual (or group) p-score and fitness; it is just that the this causal question cannot be answered by purely statistical means.

Now recall the GMA analogy, i.e. the selection / optimality links where the agents are groups. It is because G&G find a close relationship between these links holding and the absence of within-group selection, i.e. $Cov_J(p_{ij}, w_{ij}) = 0$ for all groups i, that they regard group adaptationism as intimately related to the Price approach. If one is persuaded by the alternative contextual approach, it is natural to ask what the relation is between the links holding and the absence of individual selection as defined by contextual analysis, i.e. $\beta_1 = 0$ (cf. Foster 2009).

A first step towards answering this question is to consider the relation between the absence of within-group selection and the absence of individual selection in the contextual sense. Since the Price and contextual partitions slice up the total change differently, one might think that the absence of within-group selection would be logically unrelated to the absence of individual selection in the contextual sense. Surprisingly, it turns out that this is not so.

In Appendix 2 (Proposition 1), we show that the following relation holds. For a particular p-score, if there is no within-group selection on that p-score, then either $\beta_1 = 0$ or else the p-score shows no within-group variance – in which case β_1 and β_2 are undefined. Conversely, if $\beta_1 = 0$, or if the p-score shows no within-group variance, it follows that there is no *change* due to within-group selection, i.e. the average across groups of the within-group covariance between p-score and individual fitness is zero. But this does not imply that $Cov_J(p_{ij}, w_{ij}) = 0$ for all groups i. So in short, for a given p-score, "no within-group selection" implies "no individual selection (contextual) *or* the p-score shows no within-

no within-group selection	$\forall i\ Cov_J(p_{ij}, w_{ij}) = 0$
no change due to within-group selection	$E_I[Cov_J(p_{ij}, w_{ij})] = 0$
no individual selection in the contextual sense	$\beta_1 = 0$
no within-group variance in fitness	$\forall i\ Var_J(w_{ij}) = 0$
no within-group variance in p-score	$\forall i\ Var_J(p_{ij}) = 0$

Table 3: Conditions on a single p-score

group variance", but not vice-versa.

One might conclude from this that if the absence of within-group selection characterizes group adaptationism, as G&G hold, then the absence of individual selection in the contextual sense cannot also characterize it. But this does not follow, if we adopt the 'possible' definition of scope for selection, discussed above. For surprisingly, when all possible p-scores are considered, the difference in logical strength between the conditions 'no within-group selection' and 'no individual selection (contextual)' disappears. We show this in the next section.

8 Main Results

In this section we outline our main results; full proofs are in Appendix 2. We continue to use the basic G&G model of evolution in a structured population outlined in section 3; notation remains unchanged. As before, gametic selection and mutation are assumed absent, and uncertainty is ignored. (This latter restriction could easily be relaxed.) In Table 3, we write formal definitions of the following conditions on a given p-score: "no within-group selection", "no change due to within-group selection", "no individual selection in the contextual sense", and "no within-group variance in fitness", and "no within-group variance in p-score". These conditions bear interesting logical relations to one another.

Proposition 1. *For any given p-score, the following logical implications hold:*
"no within-group variance in fitness"
$\Rightarrow$ *"no within-group selection"*
$\Rightarrow$ *"no change due to within-group selection"*
$\Leftrightarrow$ *"no individual selection (contextual)" or "no within-group variance in p-score"*

Note that the first two of these implications hold in one direction only, but the last is an equivalence.

We now consider all possible p-scores, and write formal definitions for the corresponding conditions in Table 4. Importantly, the condition "no within-group variance in any p-score" can never be satisfied, for reasons noted earlier. Similarly, the condition "no individual selection in the contextual sense on any p-score" can never be satisfied - because β_1 will be undefined for any p-score that shows no within-group variance. Note also that the condition "no within-group variance in fitness" for a single p-score, and the corresponding condition on all

no within-group selection on any p-score	$\forall p\ \forall i\ Cov_J(p_{ij}, w_{ij}) = 0$
no change due to within-group selection in any p-score	$\forall p\ E_I[Cov_J(p_{ij}, w_{ij})] = 0$
no individual selection in the contextual sense on any p-score	$\forall p\ \beta_1 = 0$
no within-group variance in fitness for any p-score	$\forall i\ Var_J(w_{ij}) = 0$
no within-group variance in any p-score	$\forall p\ \forall i\ Var_J(p_{ij}) = 0$

Table 4: Conditions on all p-scores

p-scores, are identical; since the variable 'p' does not occur in the expression '$Var_J(w_{ij})$'. (In the remainder of this section, 'all p-scores' refers to all possible p-scores, i.e. all functions from the set of individuals to $\mathbb{R}$.)

Our main result is that the following logical relations obtain between the conditions on all p-scores:

Proposition 2.
"no within-group variance in fitness"
$\Leftrightarrow$ *"no within-group selection on any p-score"*
$\Leftrightarrow$ *"no change due to within-group selection in any p-score"*
$\Leftrightarrow$ *"for each p-score, either no individual selection (contextual) or no within-group variance in that p-score"*

Note that each of these implications holds in both directions, i.e. they are equivalences. This is a striking result, given that two of the corresponding implications for a single p-score hold only in the left-right direction. To understand this, consider the first equivalence, between "no within-group variance in fitness" and "no within-group selection on any p-score". In the left-to-right direction, this is trivial. To see that it holds in right-to-left direction, suppose that fitnesses vary in at least one group. It is then possible to define a p-score which will be subject to selection within that group, simply by assigning 1 to each individual who is at least as fit as the group average, and 0 to every other individual. Therefore, the only way that there can be no within-group selection on *any* p-score is if there is no within-group variance in fitness. (This is why repression of competition, but not all groups being clonal, is sufficient for link 1 to hold.) See Figure 1 for an illustration of this point.

The second equivalence, between "no within-group selection on any p-score" and "no change due to within-group selection in any p-score", holds for essentially the same reason. Although any particular p-score can exhibit no change due to within-group selection even if it is subject to within-group selection, if all possible p-scores exhibit no change due to within-group selection this can only be because fitnesses are equal in each group, which implies the absence of within-group selection on any p-score.

Next, consider the relation between the selection / optimality links holding (on the 'possible' definition of scope for selection), and the above conditions. It is easy to see that if there is no within-group selection on any p-score, then links 1, 2 and 3 hold, where the 'agents' are groups and the objective function is group fitness (= average individual fitness). (This follows from G&G's parallel

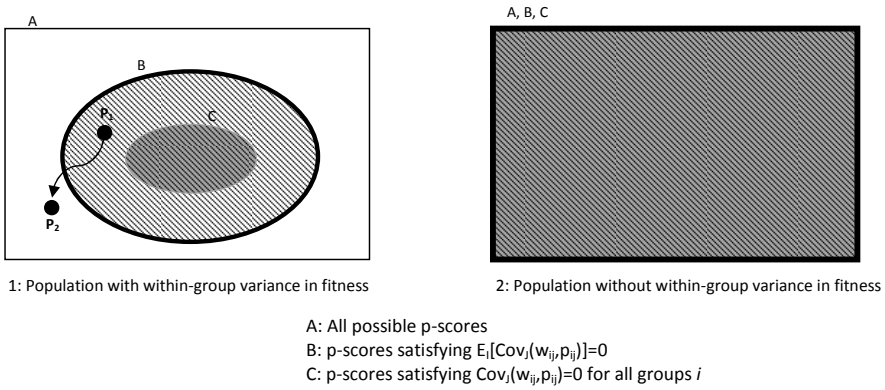

Figure 1: The left-hand box depicts a population in which there is within-group variance in fitness. The set of all possible p-scores is A. The set of p-scores for which there is no within-group selection is C. The set of p-scores for which there is no change due to within-group selection is B. Crucially, for any p-score in B but not in C, such as p_1, one can find another p-score in A but not in B, such as p_2 (see Lemma 2 in Appendix 2.) So in a population for which A=B, there can be no p-scores in B but not in C, i.e. the sets A, B, and C co-incide. Moreover, the only way in which A can equal B, is if there is no within-group variance in fitness, as in the right-hand box (see Proposition 2).

analysis in relation to all actual p-scores; see their Appendix.) We argued in section 4 that link 4 also holds in the absence of within-group selection on any p-score, and we showed in section 5 that link 5 holds in the same circumstance. Therefore, the absence of within-group selection on any p-score is sufficient for all the links to hold. Our results show that the absence of within-group selection on any p-score is equivalent to no within-group variance in fitness; so the latter condition is also sufficient for the links to hold.

What conditions are necessary for the links to hold? G&G do not explicitly discuss this; they say only that if there is within-group selection on some p-score, then the links are "not proven", which is weaker than saying that they do not hold. But the latter is in fact true. If there is within-group selection on some p-score, it is easy to show that not all of the five links can be true. In fact, something stronger can be shown, namely that either link 1, 3 or 5 must fail; see Appendix 2. So links 1, 3 and 5 jointly imply the absence of within-group selection on any p-score, which as we have seen is equivalent to the absence of within-group variance in fitness. Thus the latter condition is necessary for links 1, 3 and 5 to hold, and is thus necessary for all the links to hold.

This result is interesting, since it shows that the five links are not logically independent. For since links 1, 3, and 5 together imply the absence of within-group variance in fitness, which itself is a sufficient condition for all the links to hold – granting our argument about link 4 – it follows that links 1, 3 and 5

imply links 2 and 4. Therefore, G&G's characterization of group adaptationism, in terms of all five links holding, could in fact be re-expressed as links 1, 3 and 5 holding. This is not a criticism; to characterize a concept axiomatically one does not have to use the smallest possible axiom set; some redundancy in the axioms can be illuminating.

Therefore, granting our argument about link 4, we arrive at the following:

Proposition 3. *"links 1, 2, 3, 4 and 5 hold"*
$\Leftrightarrow$ *"links 1, 3, and 5 hold"*
$\Leftrightarrow$ *"no within-group variance in fitness"*
$\Leftrightarrow$ *"no within-group selection on any p-score"*
$\Leftrightarrow$ *"no change due to within-group selection in any p-score"*
$\Leftrightarrow$ *"for each p-score, either no individual selection (contextual) or no within-group variance in that p-score"*

These equivalences explain our claim in the previous section that the Price equation approach to multi-level selection, enshrined in equation (2), has no particular link to group adaptationism, if the latter is defined as the links holding. For while it is possible to characterize the five links holding in terms of components of equation (2), as "no change due to within-group selection in any p-score", it is equally possible to characterize their holding in terms of contextual analysis, by referring only to the parameter β_1 of equation (5). For the condition "for each p-score, either no individual selection (contextual) *or* no within-group variance in that p-score" can be re-expressed as "for each p-score, either $\beta_1 = 0$ or β_1 is undefined". So the G&G analysis of group adaptationism, under the 'possible' definition of scope for selection, provides no particular reason to favour the Price over the contextual approach to defining levels of selection.

9 G. C. Williams strikes back

We argued above that anyone accepting G.C. Williams' concept of group adaptation should distinguish between the selection / optimality links holding for the 'right' and the 'wrong' reason.The links will hold for the 'wrong' reason where there is no group functional integration and no social behaviour among individuals; but the individuals in each group happen to be equally well-adapted. This is an example of fortuitous group benefit, for Williams, and contrasts with genuine group adaptation. This means that Williams' concept of group adaptationism should be defined as 'the links hold + X'. But what is 'X'?

Intuitively, 'X' refers to the fact of group-level functional integration, e.g. the existence of a mechanism for repression of reproductive competition. The key distinction is between a case where within-group fitnesses are equalized by some such mechanism, and a case where they just happen to be equal, e.g. because the groups are clonal and individual fitness depends only on individual genotype. In the former case the individuals in each group share a common fate; in the latter case they merely have identical fates.

We introduced contextual analysis, as opposed to the Price approach, precisely to capture Williams' distinction. However the results of the last section show that, on the 'possible' definition of scope for selection, the links holding can be characterized using either the Price or the contextual approaches. So the contextual approach, as outlined above, cannot itself capture the elusive condition 'X'. So how should an advocate of Williams' concept proceed?

A natural suggestion is to modify contextual analysis by introducing a counterfactual test. Consider a simple example in which an individual's fitness depends only on a single genetic locus. There are two alleles at the locus, one of which confers a fitness advantage. All individuals in a group have the same allele, but there is between-group variance. So within-group fitnesses are equal, but not because of a group-level effect; so the links hold for the wrong reason. Consider a p-score which indicates presence or absence of the superior allele. Since this p-score shows no within-group variance, β_1 is undefined. So although individual fitness depends only on individual genotype, contextual analysis cannot detect this due to insufficient genetic variation.

A solution is to consider what would happen if genetic variation were introduced within groups. Suppose that one or more individuals had their genotype changed at the locus in question, in a way that leads to within-group variation. This will also change the fitness distribution in the population. Consider the new p-score, denoted p', that indicates presence or absence of the superior allele in this modified population. Since p' does vary within groups, β'_1 is now defined, so contextual analysis can reveal that individual fitness depends solely on individual genotype, not on group effects, i.e. $\beta'_1 \neq 0$. This shows that in the original population, the absence of within-group variance in fitness, and thus the holding of the links, was not due to a mechanism for repressing reproductive competition, but arose simply because of the absence of within-group variance in the crucial genotype.

We can generalize this example into an abstract characterization of what it means for the links to hold 'for the right reason'. Consider all the *actual* p-scores in the population. Take the subset of the actual p-scores that show no within-group variance, for which β_1 is undefined. (This subset will be non-empty in the cases that we are trying to rule out.) For each of these p-scores, we introduce within-group genetic variation in the allele that the p-score represents, by modifying the genotypes of one or more individuals. This results in a new set of actual p-scores, to which contextual analysis can be applied again, and for which the β'_1 coefficients must be well-defined. If the links hold for the 'wrong' reason, as in the example above, at least one of these β'_1 coefficients will be non-zero. If they hold for the 'right' reason, each of the coefficients will be zero, indicating that in the original population, the absence of within-group fitness variance did not arise simply because the alleles on which individual fitness depended were fixed in each group, so must have been due to a mechanism for repression of reproductive competition.

So Williams' concept of group adaptationism can be defined as 'the links holding for the right reason'. We saw above that 'the links holding' is equivalent to "for each p-score, either $\beta_1 = 0$ or β_1 is undefined". The 'right reason' can be

characterized as follows: "for all actual p-scores for which β_1 is undefined, $\beta_1' = 0$". The conjunction of these conditions thus defines group adaptationism à la Williams. This definition ensures that the distinction between group adaptation and fortuitous group benefit is respected. In a case of fortuitous group benefit, the links may hold but will not hold for the right reason, and our counterfactual test will detect this. Where the links hold for the right reason, the covariance between group p-score and group fitness, that appears in equation (2), is not simply a side effect of individual selection, but reflects a direct casual influence of group p-score on group fitness. This fits well with Williams' insistence that a group adaptation is a feature of a group that benefits it and that evolved *for that reason*.

The 'right reason' condition may seem unwieldy, referring as it does to what would happen if certain hypothetical changes were introduced into the population. It would be nicer if group adaptationism could be defined without such complications, in terms of actual statistical parameters. However it is not really surprising that this cannot be done. Williams' concept of group adaptation is explicitly causal, and it is a familiar point that causal relations cannot be fully defined in statistical terms. The distinction we have been trying to capture, between the links holding for the right and wrong reasons, is causal, and will usually be detected by contextual analysis, but not always. Where there is insufficient genetic variation for the regression coefficients of the contextual model to be defined, the distinction can only be captured by considering counterfactual scenarios.

Our proposed modification to the definition of group adaptationism – the 'right reason' condition – is expressed in terms of contextual analysis. One might think that this provides a reason to favour the contextual over the Price approach to multi-level selection. But in fact, the 'right reason' condition can be characterized using only parameters of the Price equation partition. Recall that any p-score for which β_1 is undefined must show no within-group variance, and vice-versa. We know from Proposition 1 that if a p-score does show within-group variance, then $\beta_1 = 0$ is equivalent to there being no change due to within-group selection, i.e. the second term of the Price equation (2) equals zero. Therefore, in the modified scenario, where by definition each new actual p-score does show within-group variance, the requirement that $\beta_1' = 0$ is equivalent to the absence of change due to within-group selection on the p-score. Therefore the 'right reason' condition, like the 'links holding' condition, can be equally characterized in terms of the contextual or the Price partitions.

Despite this equivalence, the contextual characterization of the 'right reason' condition is more natural. For the condition $\beta_1' = 0$ has a natural causal interpretation; it means that the gene in question does not directly affect individual fitness. By contrast, the condition 'no change due to within-group selection' has no natural interpretation. For note that, for a given p-score, this condition is *not* equivalent to the absence of within-group selection on the p-score. So one cannot capture the 'right reason' condition by requiring that there be no within-group selection in the modified population.

10 Conclusion

The idea that groups can be adaptive units is a venerable one in biology, but until Gardner and Grafen's analysis had never received a sufficiently precise formulation. Our approach has been one of critical sympathy with G&G's analysis. We have endorsed the essence of the 'formal Darwinism' project – defining adaptationism in terms of links between natural selection and optimality – and followed G&G's lead in applying this methodology to the issue of group adaptationism. Our aim has been two-fold: firstly to explore the logical consequences of the distinction between the 'actual' and 'possible' definitions of 'scope for selection'; and secondly to see whether G&G's analysis can be reconciled with G.C. Williams' influential concept of group adaptation. Our main results are the following.

On the 'actual' definition of 'scope for selection':

1. All five selection / optimality links hold for clonal groups (presuming that group fitness is a function of group phenotype alone).
2. Link 5 fails for competitively repressed groups.

On the 'possible' definition of 'scope for selection':

3. All five selection / optimality links hold for competitively repressed groups.
4. Links 1 and 3 fail for clonal groups (unless there is no within-group variance in expected fitness.)
5. The absence of within-group variance in fitness is both necessary and sufficient for the five selection / optimality links to hold.
6. Links 1, 3 and 5 are jointly equivalent to links 1, 2, 3, 4 and 5.
7. The links holding can equally be characterized in terms of the Price equation or contextual analysis.

In general:

8. Williams' concept of group adaptation can be defined as the links holding 'for the right reason'.
9. The links holding 'for the right reason' can equally be characterized in terms of the Price equation or contextual analysis.
10. Group adaptationism, in Williams' sense, cannot be fully characterized without reference to causality.

Acknowledgements

Thanks to Andy Gardner, Alan Grafen, Kevin Foster, James Marshall and Stuart West for comments and discussion. This work was supported by a research grant from the Arts and Humanities Research Council, grant no. AH/F017502/1, which we gratefully acknowledge.

A Appendix

A.1 Clonality versus repression of competition

Here we show that on the 'possible' definition of scope for selection, repression of competition suffices to prove link 1, but clonality does not.

Link 1 can be formally expressed as

$$\forall p \text{ [all groups optimal} \rightarrow \Delta\overline{p} = 0] \tag{6}$$

where '$\forall p$' means for all *possible* p-scores.

For a given p-score, it follows from equation (2) that $\Delta\overline{p} = 0$ holds if $Var_J(p_{ij}) = 0$ for all groups i (case 1), or if $Var_J(w_{ij}) = 0$ for all groups i (case 2). Focusing on case 1, we get:

$$Var_J(w_{ij}) = 0 \text{ for all groups } i \rightarrow (\text{all groups optimal} \rightarrow \Delta\overline{p} = 0) \tag{7}$$

Since (7) holds for each p-score, we can generalize to:

$$\forall p \; [Var_J(w_{ij}) = 0 \text{ for all groups } i \rightarrow (\text{all groups optimal } \rightarrow \Delta\overline{p} = 0)] \tag{8}$$

This implies:

$$\forall p \; [Var_J(w_{ij}) = 0 \text{ for all groups } i] \rightarrow \forall p \text{ [all groups optimal} \rightarrow \Delta\overline{p} = 0] \tag{9}$$

Since the condition $Var_J(w_{ij}) = 0$ for all groups i does not depend on p, equation (9) can be re-written:

$$Var_J(w_{ij}) = 0 \text{ for all groups } i \rightarrow \forall p \text{ [all groups optimal} \rightarrow \Delta\overline{p} = 0] \tag{10}$$

Equation (10) tells us that if all groups are competitively repressed, then link 1 holds on the 'possible' definition of scope for selection.

Now consider case 2. For a given p-score, we have:

$$Var_J(p_{ij}) = 0 \text{ for all groups } i \rightarrow (\text{all groups optimal} \rightarrow \Delta\overline{p} = 0) \tag{11}$$

Just as (7) leads to (9), so (11) leads to:

$$\forall p \; [Var_J(p_{ij}) = 0 \text{ for all groups } i] \rightarrow \forall p \text{ [all groups optimal} \rightarrow \Delta\overline{p} = 0] \tag{12}$$

The LHS of (12) says that no possible p-score shows within-group variance, which is impossible. So (12) is vacuously true, but does not say that link 1 holds if all groups are clonal. The latter claim is formally expressed as:

$$\forall \text{ actual } p \; [Var_J(p_{ij}) = 0 \text{ for all groups } i] \rightarrow \forall p \text{ [all groups optimal} \rightarrow \Delta\overline{p} = 0] \tag{13}$$

However (13) does not follow from (11).

So under the 'possible' definition of scope for selection, link 1 holds for competitively repressed groups but not for clonal groups.

A.2 Main results

Proposition 1. *For a given p-score:*
$\forall i\ Var_J(w_{ij}) = 0$
$\Rightarrow \forall i\ Cov_J(p_{ij}, w_{ij}) = 0$
$\Rightarrow E_I[Cov_J(p_{ij}, w_{ij})] = 0$
$\Leftrightarrow \beta_1 = 0$ *or* $\forall i\ Var_J(p_{ij}) = 0$

Proof. The first two implications are trivial, so we just prove the last equivalence.
Recall the Price partition:

$$\overline{w}\Delta\overline{p} = Cov_I(w_i, p_i) + E_I[Cov_J(p_{ij}, w_{ij})] \tag{2}$$

And the contextual one :

$$\overline{w}\Delta\overline{p} = \beta_2 Var_I(p_i) + \beta_1 Var_{I\text{x}J}(p_{ij}) \tag{13}$$

1. In the right-to-left direction:
Suppose that either $\beta_1 = 0$ or $\forall i\ Var_J(p_{ij}) = 0$.

- If $\forall i\ Var_J(p_{ij}) = 0$, then both terms of the Price partition are zero, as are both terms of the contextual partition.
- If $\beta_1 = 0$, then from the contextual partition we have $\overline{w}\Delta\overline{p} = \beta_2 Var_I(p_i)$. Now consider the group term of the Price equation, $Cov_I(p_i, w_i)$. This always equals $Cov_{I\text{x}J}(p_i, w_{ij})$. Substituting for w_{ij} using regression equation $w_{ij} = \beta_1 p_{ij} + \beta_2 p_i + e_{ij}$, we get $Cov_{I\text{x}J}(p_i, w_{ij}) = \beta_2 Cov_I(p_i, p_i) = \beta_2 Var_I(p_i)$. Therefore $Cov_I(p_i, w_i) = \beta_2 Var_I(p_i) = \overline{w}\Delta\overline{p}$, so $E_I[Cov_J(p_{ij}, w_{ij})] = 0$.

2. In the left-to-right direction:
Suppose that $E_I[Cov_J(p_{ij}, w_{ij})] = 0$. Therefore, $\overline{w}\Delta\overline{p} = Cov_I(p_i, w_i) = Cov_{I\text{x}J}(p_i, w_{ij})$.
Again, substitute in the regression equation $w_{ij} = \beta_1 p_{ij} + \beta_2 p_i + e_{ij}$, which gives:
$\overline{w}\Delta\overline{p} = Cov_{I\text{x}J}(p_i, w_{ij}) = \beta_1 Var_I(p_i) + \beta_2 Var_I(p_i)$.
But the contextual partition says that:
$\overline{w}\Delta\overline{p} = \beta_2 Var_I(p_i) + \beta_1 Var_{I\text{x}J}(p_{ij})$.
Therefore, $\beta_1 Var_I(p_i) = \beta_1 Var_{I\text{x}J}(p_{ij})$, which implies that either $\beta_1 = 0$, or else $Var_I(p_i) = Var_{I\text{x}J}(p_{ij})$.
If $Var_I(p_i) = Var_{I\text{x}J}(p_{ij})$, then the total variance in p-score equals the between-group variance, which implies that $\forall i\ Var_J(p_{ij}) = 0$.

□

Lemma 1. $\forall i\ Var_J(w_{ij}) = 0 \Leftrightarrow \forall p\ \forall i\ Cov_J(p_{ij}, w_{ij}) = 0$

Proof. The left-to-right direction is trivial. In the right-to-left direction, suppose that $\forall p\ \forall i\ Cov_J(p_{ij}, w_{ij}) = 0$ but that there is some within-group variance in fitness, i.e. $\exists i\ Var_J(w_{ij}) \neq 0$. Assume without loss of generality that there is only one group g in which fitnesses vary. Now consider the p-score p' defined as

follows: $p'_{gj} = 1$ for all individuals j in group g such that $w_{gj} > w_g$, and $p'_{gj} = 0$ for all other individuals in g. By construction, in group g, $Cov_J(p_{gj}, w_{gj}) > 0$, which contradicts the hypothesis.

□

Lemma 2. $\forall p\ \forall i\ Cov_J(w_{ij}, p_{ij}) = 0 \Leftrightarrow \forall p\ E_I[Cov_J(w_{ij}, p_{ij})] = 0$

Proof. The left-to-right direction is obvious: the average of components equal to zero is zero. In the right-to-left direction, suppose that $\forall$ p-score $E_I[Cov_J(w_{ij}, p_{ij})] = 0$, but that there exists a p-score p and a group g such that $Cov_J(w_{gj}, p_{gj})] > 0$. As the non-zero covariances must average out, it follows that there exists a group g' such that $Cov_J(w_{g'j}, p_{g'j}) < 0$.

Assume without loss of generality that g and g' are the only groups in which the covariances are not zero. Recall that a p-score is just a function $I \text{x} J \to \mathbb{R}$. Let a new p-score p' be defined as follows: $\forall i \neq g'\ \forall j\ p'_{ij} = p_{ij}$, and $\forall j\ p'_{g'j} = 1 - p_{g'j}$. Informally put, p' is identical to p in general, except in the groups for which the covariance is negative, where they are opposite.

So $\forall i \neq g'\ Cov_J(w_{ij}, p'_{ij}) = Cov_J(w_{ij}, p_{ij}) \geq 0$; but $Cov_J(w_{g'j}, p'_{g'j}) = -Cov_J(w_{g'j}, p_{g'j}) > 0$.

Overall, $\forall i\ Cov_J(w_{ij}, p'_{ij}) \geq 0$, and the inequality is strict for groups g and g'. As a consequence, $E_I[Cov_J(w_{ij}, p'_{ij})] > 0$, which contradicts the hypothesis.

□

Proposition 2.

$\forall i\ Var_J(w_{ij}) = 0$
$\quad \Leftrightarrow \forall p\ \forall i\ Cov_J(p_{ij}, w_{ij}) = 0$
$\quad\quad \Leftrightarrow \forall p\ E_I[Cov_J(p_{ij}, w_{ij})] = 0$
$\quad\quad\quad \Leftrightarrow \forall p\ [\beta_1 = 0\ \textit{or}\ \forall i\ Var_J(p_{ij}) = 0]$

Proof. The first equivalence corresponds to Lemma 1; the second equivalence to Lemma 2. The third one is trivially obtained from the final equivalence of Proposition 1 (if an equivalence holds for any given p, then it holds for all p). □

Proposition 3.

"links 1, 2, 3, 4 and 5 hold"
$\quad \Leftrightarrow$ *"links 1, 3 and 5 hold"*
$\quad\quad \Leftrightarrow \forall i\ Var_J(w_{ij}) = 0$
$\quad\quad\quad \Leftrightarrow \forall p\ \forall i\ Cov_J(p_{ij}, w_{ij}) = 0$
$\quad\quad\quad\quad \Leftrightarrow \forall p\ E_I[Cov_J(p_{ij}, w_{ij})] = 0$
$\quad\quad\quad\quad\quad \Leftrightarrow \forall p\ [\beta_1 = 0\ \textit{or}\ \forall i\ Var_J(p_{ij}) = 0]$

Proof. The last three equivalences correspond to Proposition 2. Gardner and Grafen (2009, pp. 12–13) have already proved that $\forall p\ \forall i\ Cov_J(p_{ij}, w_{ij}) = 0$ implies links 1-5, which trivially imply links 1, 3 and 5. Thus we only have to show that links 1, 3 and 5 jointly imply one of the last four conditions above.

To show that links 1, 3 and 5 jointly imply $\forall i\ Var_J(w_{ij}) = 0$, suppose that there is some group i such that $Var_J(w_{ij}) \neq 0$. There are three possible cases:

- all groups are optimal. Then, for any p-score, there will be no change due to between-group selection. However, because within-group fitnesses are not identical in group i, we can find a p-score for which $E_I[Cov_J(p_{ij}, w_{ij})] \neq 0$, by Lemmas 1 and 2. Therefore there is scope for selection, so link 1 is false.

- all groups are equally suboptimal. By the same reasoning as in the previous case, there is scope for selection, so link 3 is false.

- groups vary in optimality. Since within-group fitnesses are not identical in group i, we can find a p-score for which $E_I[Cov_J(p_{ij}, w_{ij})] \neq 0$. Therefore, the change in this p-score, $\Delta\overline{p}$ is not equal to $Cov_I(p_i, w_i)$, so link 5 is false.

Therefore, if it's false that $\forall i \; Var_J(w_{ij}) = 0$, then either link 1, 3 or 5 is false.

□

References

[1] Buss, L. *The Evolution of Individuality*, Princeton University Press, Princeton NJ, 1987.

[2] Foster, K. R. 'A defense of sociobiology', *Cold Spring Harbour Symposium in Quantitative Biology* 74: 403–418, 2009.

[3] Frank, S.A. 'Repression of competition and the evolution of cooperation', *Evolution* 57: 693–705, 2003.

[4] Gardner, A. & A. Grafen, 'Capturing the superorganism: a formal theory of group adaptation', *Journal of Evolutionary Biology*, 22:659–671, 2009.

[5] Goodnight, C.J., J.M. Schwartz & L. Stevens, 'Contextual analysis of models of group selection, soft selection, hard selection, and the evolution of altruism', *American Naturalist*, 140:743–761, 1992.

[6] Grafen, A. 'Natural selection, kin selection and group selection', in J.R. Krebs & N.B. Davies, eds., *Behavioural Ecology, 2nd edition*, 62–84, Blackwell Scientific Publications, Oxford, 1984.

[7] Grafen, A. 'A geometric view of relatedness', *Oxford Surveys in Evolutionary Biology*, 2:28–89, 1985.

[8] Grafen, A. 'A first formal link between the Price equation and an optimization program', *Journal of Theoretical Biology*, 217:75–91, 2002.

[9] Grafen, A. 'Optimization of inclusive fitness', *Journal of Theoretical Biology*, 238:541–563, 2006.

[10] Grafen, A. 'The simplest formal argument for fitness optimization', *Journal of Genetics*, 87(4):421–433, 2008.

[11] Heisler, I. L. & J. Damuth, 'A method for analyzing selection in hierarchically structured populations', *American Naturalist*, 130:582–602, 1987.

[12] Hölldobler, B. & E.O. Wilson, *The Superorganism: the Beauty, Elegance and Strangeness of Insect Societies*, W.W. Norton, London, 2009.

[13] Maynard Smith, J. & E. Szathmary, *The Major Transitions in Evolution*, Oxford University Press, Oxford, 1995.

[14] Michod, R. E., *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality*, Princeton University Press, Princeton NJ, 1999.

[15] Nunney, L. 'Group selection, altruism, and structured-deme models', *American Naturalist*, 126: 212–230, 1985.

[16] Okasha, S. 'Multi-level selection and the partitioning of covariance: a comparison of three approaches', *Evolution* 58: 486–494, 2004.

[17] Okasha, S. *Evolution and the Levels of Selection*, Oxford University Press, New York, 2006.

[18] Price, G. R. 'Extension of covariance selection mathematics', *Annals of Human Genetics* 35: 485–490, 1972.

[19] Queller, D. C. & J. E. Strassmann, 'Beyond society: the evolution of organismality', *Philosophical Transactions of the Royal Society B* 364: 3143–3155, 2009.

[20] Ratnieks, F. L. W. & H. K. Reeve 'Conflict in single-queen hymenopteran societies: the structure of conflict and processes that reduce conflict in advanced eusocial species', *Journal of Theoretical Biology* 158: 33–65, 1992.

[21] Seeley, T.D. 'The honey bee colony as a superorganism', *American Scientist* 77: 546–553, 1989.

[22] Seeley, T.D. 'Honey bee colonies are group-level adaptive units', *American Naturalist* 150: S22–S41, 1997.

[23] Sober, E. & D.S. Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behaviour*, Harvard University Press, Cambridge MA, 1998.

[24] Sober, E. & D.S. Wilson, 'Adaptation and Natural Selection Revisited', *Journal of Evolutionary Biology*, 24:462–468, 2011.

[25] Williams, G.C. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*, Princeton University Press, Princeton, 1966.